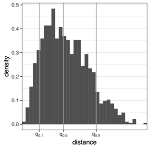
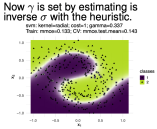
Introduction to Machine Learning

SVM Model Selection

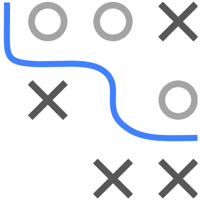




Learning goals

- Know that the SVM is sensitive to hyperparameter choices
- Understand the effect of different (kernel) hyperparameters

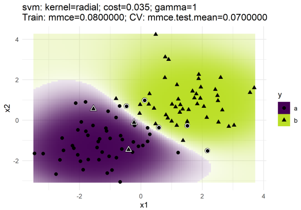
MODEL SELECTION FOR KERNEL SVMS

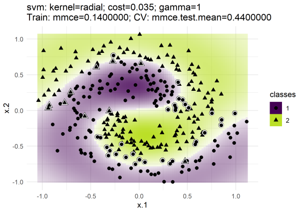


- “Kernelizing” a linear algorithm effectively turns this algorithm into a family of algorithms — one for each kernel. There are infinitely many kernels, and many efficiently computable kernels.
- However, the choice of C, the choice of the kernel, the kernel parameters are all up to the user.
- On the one hand this allows very flexible modelling, and also to incorporate prior knowledge into the learning process.
- On the other hand this puts a huge burden on the user. The machine has no mechanism for identifying a good kernel by itself.
- SVMs are somewhat sensitive to its hyperparameters and should always be tuned.
- Gaussian processes are very related kernel methods, with the big advantage that kernel parameters are directly estimated during training.

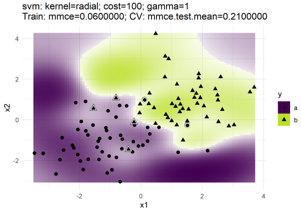
SVM HYPERPARAMETERS

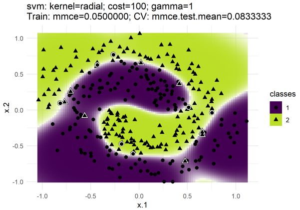
Small C “allows” for margin-violating points in favor of a large margin.





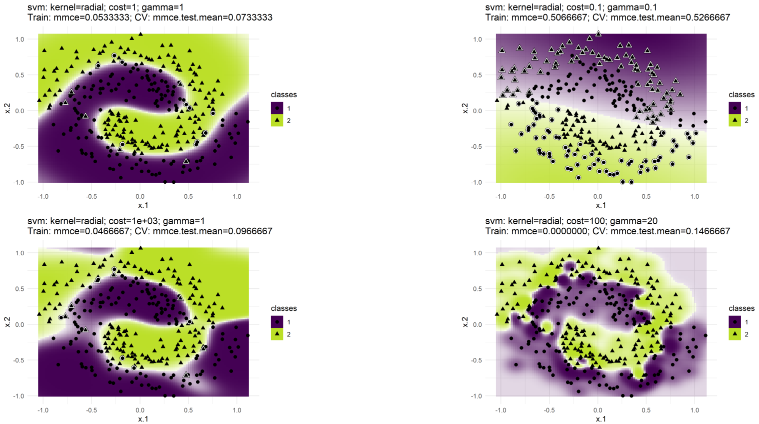
Large C penalizes margin violators, decision boundary is more “wiggly”.





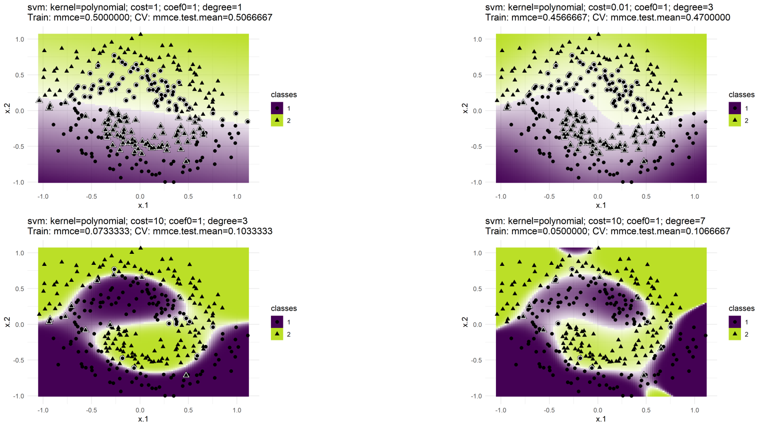
SVM HYPERPARAMETERS

Hyperparameters strongly influence the model: RBF kernel.



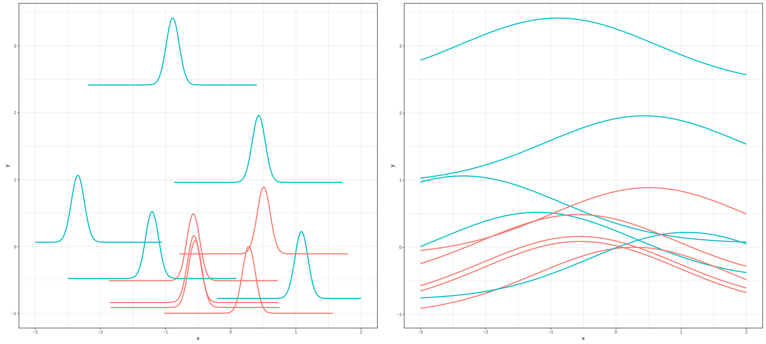
SVM HYPERPARAMETERS

Hyperparameters strongly influence the model: Polynomial kernel.



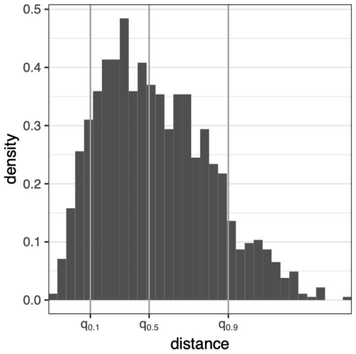
RBF SIGMA HEURISTIC

For the RBF kernel $k(\mathbf{x}, \tilde{\mathbf{x}}) = \exp(-\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|^2}{2\sigma^2})$ a simple heuristic exists for the width hyperparameter σ^2.

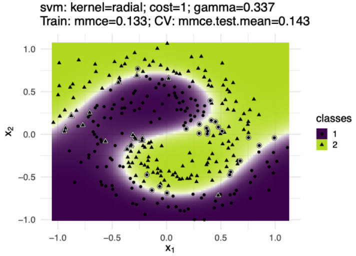


RBF SIGMA HEURISTIC

- Draw a random subset of the data.
- Compute pairwise distances $\|\mathbf{x} - \tilde{\mathbf{x}}\|$.
- Take a "central quantile" from their distribution, e.g., the median.
- This relates the kernel width to the "average distance" between points, which does make intuitive sense.



Now γ is set by estimating is inverse σ with the heuristic.



SVM HYPERPARAMETERS

- RBF-SVM parameters are often optimized on log-scale, as we want to explore large values and values close to 0.
- E.g.: $C \in [2^{-15}, 2^{15}], \gamma \in [2^{-15}, 2^{15}]$
- The cross-validated performance landscape often forms a characteristic "ridge" with a larger area of equally good values.